

81 J. Biol. Chem. 274:26015-26020(1999).
 82 [2]
 83 SOURCE: OF 19-44.
 84 TISSUE: MILK;
 85 MEDLINE: 96139344; PubMed 8547333;
 86 POLJANOV S.N., POLJANOV T.E., NEXO E.;
 87 "Transcobalamin from cow milk: isolation and physico-chemical
 88 properties";
 89 Biochim. Biophys. Acta 1292:113-119(1996).
 90 FUNCTION: Primary vitamin B12-binding and transport protein.
 91 Delivers cobalamin to cells.
 92 SUBCELLULAR LOCATION: Secreted.
 93 TISSUE SPECIFICITY: EXPRESSED IN MAMMARY GLAND, KIDNEY, LYMPHATIC
 94 NODES AND LIVER.
 95 SIMILARITY: BELONGS TO THE EUKARYOTIC COBALAMIN TRANSPORT PROTEINS
 96 FAMILY.
 97
 98 This SWISS-PROT entry is copyright. It is produced through a collaboration
 99 between the Swiss Institute of Bioinformatics and the EMBL outstation
 100 the European Bioinformatics Institute. There are no restrictions on its
 101 use by non-profit institutions as long as its content is in no way
 102 modified and this statement is not removed. Usage by and for commercial
 103 entities requires a license agreement. (See <http://www.isb-sib.ch/anonymous/>
 104 or send an email to license@isb-sib.ch).
 105
 106 EMBL: AF121289; AAC23829.1;
 107 InterPro: IPR002157; Cobalamin_bind.
 108 Pfam: PF01122; Cobalamin_bind. 1.
 109 PROSITE: PS00468; COBALAMIN_BINDING. 1.
 110 TRANSPORT: Cobalt transport; Signal.
 111 SIGNAL 1 18
 112 CHAIN 19 432 TRANS-COBALAMIN 11.
 113 DISULFID 21 270
 114 DISULFID 116 312
 115 DISULFID 165 208
 116 CARBOHYD 94 94 N-LINKED (GLUCNAc...) (POTENTIAL).
 117 SEQUENCE 432 AA: 43 676 MW: 50.441kDa pI: 5.65
 118
 119 Query Match 68.5% Score 1423 DB 1: Length 432;
 120 Best Local Similarity 63.0% Pred. No. 1,3e-194;
 121 Matches 272; Conserved 35; Mismatch 69; Indels 56; Gaps 2.
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

DB 421 DQITFLRWVW 432
 RESULT 5
 ID COI_HUMAN STANDARD; PRI: 433 AA.
 AC P20061;
 DT 01-FEB-1991 (Rel. 17, created)
 DT 01-FEB-1991 (Rel. 17, last sequence update)
 DT 01-MAR-2002 (Rel. 41, last annotation update)
 DE Transcobalamin 1 precursor (Trc1) (Trc1).
 GN Trc1 OR Trc1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catartia; Hominoidea; Homo.
 OX MGI_Taxid 9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 89380156; PubMed: 277761;
 RA Johnston J., Pollekens J., Allen R.H., Beetham N.;
 RT "Structure of the cDNA encoding transcobalamin 1, a neutrophil
 R1 granule protein";
 R1 J Biol Chem 264:15754-15757(1989).
 CC -1: FUNCTION: VITAMIN B12 BINDING PROTEIN. TRANSPORTS COBALAMIN INTO
 CC CELLS.
 CC -1: SUBCELLULAR LOCATION: Secreted.
 CC -1: TISSUE SPECIFICITY: NEUTROPHIL GRANULE PROTEIN.
 CC -1: PTM: CONTAINS ABOUT 30% CARBOHYDRATES.
 CC -1: SIMILARITY: BELONGS TO THE EUKARYOTIC COBALAMIN TRANSPORT PROTEINS
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J05068; AAA61058.1;
 CC PIR: A34227; A34227.
 CC MIM: 129905;
 CC InterPro: IPR002157; Cobalamin_bind.
 CC Pfam: PF01122; Cobalamin_bind. 1.
 CC PROSITE: PS00468; COBALAMIN_BINDING. 1.
 CC TRANSPORT: Cobalt transport; glycoprotein; Signal.
 CC SIGNAL 1 24
 CC CHAIN 24 433 TRANS-COBALAMIN 1.
 CC DISULFID 26 265 BY SIMILARITY.
 CC DISULFID 105 308 BY SIMILARITY.
 CC DISULFID 155 197 BY SIMILARITY.
 CC CARBOHYD 160 160 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 216 216 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 316 316 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 337 337 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 343 343 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 349 349 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 354 354 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC CARBOHYD 369 369 N-LINKED (GLUCNAc...) (POTENTIAL).
 CC SEQUENCE 433 AA: 46194 MW: 561867914 Da pI: 5.654;
 4 LGATFLVGLVGAITGCEIPEDMSHLVERLQHLIPWMDRLSLRHN-----PSIYGL 58
 10 VGLLFST-IPSDULRLEVS-FNYT-----KLPFLNIMVSYNNGTSANVAVLSL 61
 59 PISNGATKEDYVHSKIKGVGGCTIGSNAPSTWVWQGRKSMGALATLAL----- 112
 62 KLVGIGVIGTLMGKKIKQDTKYNVNSPL-----SIWSSDEALITLALGVCRNA 108

0Y	113		PANCFEYBOKRQKPLVSQ		-----LK	132
1b	109	HENILTYOYH TGGT LRFEGAE LNENFAHTTTI INHYGALVATCI NGRSLAELVN			-----	168
0Y	143	WLEDEKR-----ADTAAAGLAPTCOL KRSNN-----DGRQRITMALIKVREEL			-----	179
1b	169	HETPFKNYYVFSSSEVFSGAAVALALGVKEFI INGLFAGOGELINNYIKELVKLI			-----	228
0Y	190	LKATPECHPCNYYSTHALGPI MESTMPRAFTGTAC KAVALALASGCALAMLRIC			-----	239
1b	229	I SEPPFNH I PRTDTTPZMAI IVSQQVYDFQAVYGF PVVI SFLSDZADGNMADA			-----	289
0Y	240	GILPVLNHRKYTHL-----IPPGCLAPVMLEPAHEI PGCGELISTGLCV			-----	285
1b	299	CATFALMCPPTCTHHGSSQSVASCNHLSAD ... PRVVIVHFGZYVAYVY			-----	349
0Y	299	LEI PFYPSNGVAAGTVAFI PFAHLTG ... VTTLIAATGQVTVGVVAAA			-----	399
1b	341	RINEIYATVYNLVINSVFI SVFRAGMMGNTI TPMPHDS--WRHYTCTGTCANNR			-----	397
0Y	449	FPRPGCISRGSTTI GGLAIYFFPGCTTIF			-----	472
1b	498	DRAIVEILL-SGSPHSQAAGSVAVENCPNF EVD			-----	499
<hr/>						
RESULT	6					
IF_MOUSE						
ID	IF_MOUSE	STANDARD:	PROT	417	AA	
AC	P57878.					
DI	01-OCT-1996 (rel. 34, Created)					
DI	01-OCT-1996 (rel. 34, Last sequence update)					
DI	01-MAR-2002 (rel. 41, Last annotation update)					
DE	Intestinal factor (proliferative vit) (cancer inhibitory factor)					
ON	GIF.					
OS	Mus musculus (Mouse)					
OC	Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
OX	MRI_TaxId=10090;					
ON	[]					
KD	SEQUENCE FROM N.A.					
RC	STRAIN-HALB/C; HISSOF Stomach;					
RX	MEDLINE=94075147; PubMed=8253786;					
RA	Lorenz R.G., Gordon J.L.;					
RT	"Use of transgenic mice to study regulation of gene expression in the					
RL	patent cell lineage of gastric units."					
CC	J. Biol. Chem. 268:26559-26570(1993).					
CC	-+ PROMOTER PROMOTES ABSORPTION OF THE ESSENTIAL VITAMIN COBALAMIN					
CC	(CRU) IN THE ILEUM BY SPECIFIC RECEPTOR-MEDIATED ENDOCYTOSIS.					
CC	-+ SIMILARITY PRIORS TO THE EPIDEMIOLOGIC COBALAMIN TRANSFER PROTEIN					
CC	FAMILY					
CC	-----					
CC	This SWISS Prot entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL database					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed, usage by and for commercial					
CC	entities is prohibited without prior written permission. ***set up by swiss					
CC	or send an email to license@sib.ch).					
CC	-----					
DR	EMBL: L24191; AAA37882.1; -					
DR	EMBL: L24192; AAA37881.1; -					
DR	MDG: MG1:1202494; GI1.					
DR	InterPro: IPRO02157; CdbAlamr_Bind.					
DR	Pham: PH01122; CdbAlamr_Bind.					
DR	ProSite: PS00468; COBALAMIN_BINDING; 1					
FW	Transports cobalamin transport glycoprotein signal					
FT	SIGNAL 1 18 POTENTIAL					
FT	CHAIN 19 417 INTRINSIC FACTOR					
FT	DISULFID 26 246 BY SIMILARITY					
FT	DISULFID 103 268 BY SIMILARITY					
FT	DISULFID 143 162 BY SIMILARITY					

[illegible]

FT	DLSUFLID	155	197		BY SIMILARITY.
FT	CABP_HYD	86	89	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	160	160	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	165	165	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	178	178	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	411	413	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	427	427	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	444	444	N-LINKED (GLUCNA-)	(POTENTIAL).
FT	CABP_HYD	359	359	N-LINKED (GLUCNA-)	(POTENTIAL).
SJ	SEQUENCE	416 AA:	4696 MS:	38000 EFCECHIAAEF CRG64:	
<hr/>					
Query Match					
Post Local Similarity:		9.5%	Score 184.5:	DB 1:	Length 416:
Matches 101:		Conserved: 42:	Mismatches 119:	Indels 165:	Gaps
<hr/>					
QY	4	LCAPLPTGVCAITKMPVPEEDM-SHLYERLQHLLIMMERISEHL-----NPSIVGL	58		
Db	9	VGLLETLTIPSDQSIVSKRSHI-----PIISAMD--NI FQIRGTYSASTLSQ	60		
QY	59	RISSQ AETPRVYKHSLRGSTLTGSASFSELDGGGRISM GELATYIALRA-	114		
Db	61	KIASIQHSLEHL-----SGRKQDMKRMRNSNTSQAALPLAGAC	105		
QY	115	---NCFVRGH			
Db	100	KLPDVERLHDHVLVALDKKKREELNNMLHNSNPENINYQLSVTHLDFEKNYSIN	165		
QY	131	LKEELLETER-----FAITAMAAVLAIGKRNH-----	161		
Db	164	VHPHFNFHRNLNLSHQVTVGVAVVAALVPFSISQRLKAATKRYSTGYHSIV	225		
QY	172	FTEAETLVGEETVEETFA--LGGWVDTDA--LTPSTPMAIGTAY*	218		
Db	226	HRTSPRVVVSIFPLADKPTQPS-----SDETRPKRMQAKVTPQPT	271		
QY	219	ARKALLASGGDNVQNALIKSQLPVLRKYTDLTHTSLAIRKYLEMAATINQDEL	278		
Db	272	---HSGVVPEFLPAAGCTIHAIQPYVD-VKRTLVVAVVNTPEPV----	418		
QY	279	LVTHLVETLPPVQSTNVSLACTIVE-----GVYKAFI-----GMYET	221		
Db	319	---VVTTLS--ENVIVTYQFINLSMDITIVLTVVMALALNRITTYQWTF	470		
QY	322	QASLSQVLTYSWGKAA--GREPWQ	345		
Db	371	P---WQVETVSCVWANKHSETWE	393		
<hr/>					
RESULT 10					
LIPA_CHLPN STANDARD; PRT; 307 AA.					
ID	LIPA_CHLPN	Q9JQ18:			
AC	G9Z7744: Q9JQ18:	30-MAY-2000 (rel. 39, Created)			
DI	30-MAY-2000 (rel. 39,	Last sequence update)			
DI	16-OCT-2001 (rel. 40,	Last annotation update)			
DE	Elliptic acid synthetase (LIP SYN) (catalytic synthase).				
GN	LIPA op GPM083 OR CP1038.				
OS	Chlamydia pneumoniae (Chlamydiophila pneumoniae).				
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiophila.				
QC	NCBI Taxid 83588;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-CMU029;				
RX	KCMT-HB-99204E06; PubMed 10192488;				
PA	Kilmar S, Mitchell W, Harris E, Lamb J, Fan Z, Hyman P.W.,				
RT	Olinier L, Grimwood J, Davis R.W, Stephens R.S.G.				
RL	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."				
SI	Nat. Genet. 21:385-389(1999).				
SP	SEQUENCE FROM N.A.				
XC	STRAIN AR39;				
XZ	MEDLINE-20150265; PubMed 1094945;				

RESULT 12

GNL_YEAST STANDARD PRT: 2672 AA.

AC P41892

BT 01-FEB-1994 (Rel. 28, Created)

BT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Translational activator GCN1

GN GCN1 OR YG1196W OR YG119

OS Saccharomyces cerevisiae (Baker's yeast).

PC Pharyngota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes:

OC Saccharomycetales: Saccharomycetaceae: Saccharomycetes.

EX NBI_TaxId:4932;

GN NBI_TaxId:4932;

RP SEQUENCE FROM N.A.

RX MEDLINE:9326194; PubMed 8497269;

RA Martin M.J., Crook D., Hinnenrich A.G.

RT "GCN1, a translational activator of GCN4 in Saccharomyces cerevisiae, is required for phosphorylation of eukaryotic translation initiation factor 2 by protein kinase GCN2."

RT Mol. Cell. Biol. 13:3541-3550(1993)

RL [2]

RN SEQUENCE FROM N.A.

RP STRAIN-S288c / Y1679;

RX MEDLINE:9719797; PubMed-9646087;

RA Coglievina M., Klima R., Hertant I., Deliceti D., Zaccaria P., Hirschi C.V.

RT "Sequencing of a 40.5 kb fragment located on the left arm of chromosome VII from Saccharomyces cerevisiae."

RL Yeast 13:55-64(1997).

CC -1- FUNCTION: TRANSLATIONAL ACTIVATOR OF GCN4. MAY BE INVOLVED IN IN AMINO ACID STAVED CELLS. REQUIRED IN VIVO FOR THE PHOSPHORYLATION OF EIF-2-ALPHA ON SERINE-52 BY THE PROTEIN KINASE GCN2

CC -1- SUBUNIT: COMPONENT OF A HETEROMERIC COMPLEX THAT INCLUDES GCN1 AND GCN20.

CC -1- SIMILARITY: STRONG, TO S. POMBE SPAC1066.05C.

CC -1- SIMILARITY: CONTAINS 20 HEAT REPEATS.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement. (See http://www.ebi.ac.uk/seqdb/doc/ or send an email to license@sdb.ebi.ac.uk).

CC -----

DR EMBL: L12467; AAA14635.1;

DR EMBL: X91837; CAA62949.1;

DR EMBL: Z72777; CAA60937.1;

DR PIR: A48126; A48126.

DR SDB: S000316; GCN1.

DR InterPro: IPR000357; HEAT_repeat.

DR PROSITE: PS00077; HEAT_REPEAT_4.

DR Translational activator GCN1; repeat.

FT REPEAT 932 970 HEAT 1.

FT REPEAT 1030 1067 HEAT 2.

FT REPEAT 1099 1138 HEAT 3.

FT REPEAT 1243 1281 HEAT 4.

FT REPEAT 1284 1321 HEAT 5.

FT REPEAT 1405 1442 HEAT 6.

FT REPEAT 1484 1521 HEAT 7.

FT REPEAT 1523 1559 HEAT 8.

FT REPEAT 1561 1598 HEAT 9.

FT REPEAT 1603 1640 HEAT 10.

FT REPEAT 1641 1679 HEAT 11.

FT REPEAT 1721 1758 HEAT 12.

FT REPEAT 1760 1796 HEAT 13.

FT REPEAT 1862 1903 HEAT 14.

FT REPEAT 1905 1942 HEAT 15.

FT REPEAT 1947 1984 HEAT 16.

FT REPEAT 1995 2024 HEAT 17.

RESULT 13

SL55_RAT STANDARD PRT: 619 AA.

AC G030087

BT 16-OCT-2001 (Rel. 40, Created)

BT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sodium iodide cotransporter (Na/I) (Solute symporter) (Sodium iodide symporter) (Na/I symporter).

GN SLC5A5 OR NIS.

OS Rattus norvegicus (rat).

PC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:

OC Mammalia: Eulipotyria: Rodentia: Sturniinae: Muridae: Rattus.

EX NBI_TaxId:10116;

GN NBI_TaxId:10116;

RP SEQUENCE FROM N.A.

RX TISSUE=Thyroid;

RX MEDLINE:96158880; PubMed 8579252;

RA Dai G., Levy O., Carrasco N.

RT "Cloning and characterization of the thyroid iodide transporter."

RL Nature 379:458-460(1996).

CC -1- FUNCTION: MEDIATES IODIDE UPTAKE IN THE THYROID GLAND.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE SODIUM-DEPENDENT SYMPORTER FAMILY (SDF).

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC

FT REPEAT 2097 2134 HEAT 18.

FT REPEAT 2290 2328 HEAT 19.

FT REPEAT 2312 2384 HEAT 20.

SO SEQUENCE 2672 AA; 296693 MW; 980FD00375459D10 CRC64;

Query Match 5.0% Score 97, DN 12, Length 2672;

Best local similarity 22.0% Prod. No. 8.6;

Matches 102, Coverage 61, Mismatch 179, Indels 138, Gaps 20;

57 GAKUSHA GINADLDS LAKYVQL SAS 90

1461 GVESSQAAAEVLSGLKLEMLPELVAVREAVPEGRHLLFVGVNFA 1720

91 EKKCGGCKRMQALYL ALPQVVGKRG---PEVSGK--- 132

1721VYMLLEFVLSALAGNHLALALALVAVNATVAVVLLFELFG 1771

133 WPIPKKPAIDPFAAMG-LAVTQI---KSNFN-----PEPKRIT 169

1772 MFGNDPFAISVGLFQELFQVTLSSNHLSEHQNDSEKIVTVLQDEPRHIL 1831

170 MATTVKPP---TKKQTPKPKGVVTVDAVPTMSMDS---ALHSTAGYAP 220

1842 ALIVKPKGSGVPAIVTVINPAIVINPAVPTLPI-TGMIVTIAGSSNVPNTA 1860

221 VALIASLQDAFCAHLISGLLVCHKCYLILFDCAPRWLAAV ETITQ 274

1891 AQLIGLVKRVGNAL---SLKSLP-LSLLELSNLSNGVCIATLHLSASTETISQ 1947

275 TQPTISVTVGVSTLTPTQSSVLCAGVEVEVFAHHCSTFTVQASNGEPLTSM 334

1948 FQ-----STIVNITPAILDESATVEAAALSFDPQIVV 1982

335 GRAGGPFVWLR-----IPNTELLGLADVAKKGGIT 369

1993 GRTAVLEVLYLHLLKESRDFALGQETMSKSDVI 2022

F1	DMAIN	1	16	EXTRECELLULAR (POTENTIAL).
F1	TRANSSEM	17	47	POTENTIAL.
F1	DMAIN	38	54	CYTOPLASMIC (POTENTIAL).
F1	TRANSSEM	54	74	POTENTIAL.
F1	DMAIN	75	88	EXTRACELLULAR (POTENTIAL).
F1	TRANSSEM	89	109	POTENTIAL.
F1	DMAIN	110	136	CYTOPLASMIC (POTENTIAL).
F1	TRANSSEM	137	157	POTENTIAL.
F1	DMAIN	158	163	EXTRACELLULAR (POTENTIAL).
F1	TRANSSEM	164	184	POTENTIAL.
F1	DMAIN	185	186	CYTOPLASMIC (POTENTIAL).
F1	TRANSSEM	187	207	POTENTIAL.
F1	DMAIN	208	241	EXTRACELLULAR (POTENTIAL).
F1	TRANSSEM	242	262	POTENTIAL.
F1	DMAIN	263	286	CYTOPLASMIC (POTENTIAL).
F1	TRANSSEM	287	307	POTENTIAL.
F1	DMAIN	308	326	EXTRACELLULAR (POTENTIAL).
F1	TRANSSEM	327	347	POTENTIAL.
F1	DMAIN	348	391	CYTOPLASMIC (POTENTIAL).
F1	TRANSSEM	392	412	POTENTIAL.
F1	DMAIN	413	416	EXTRACELLULAR (POTENTIAL).
F1	TRANSSEM	417	437	POTENTIAL.
F1	DMAIN	438	444	CYTOPLASMIC (POTENTIAL).
F1	TRANSSEM	445	465	POTENTIAL.
F1	DMAIN	466	520	EXTRACELLULAR (POTENTIAL).
F1	TRANSSEM	521	541	POTENTIAL.
F1	DMAIN	542	618	CYTOPLASMIC (POTENTIAL).
F1	MOD RES	551	551	PHOSPHORYLATION (BY GAPR) (POTENTIAL).
F1	CARBOHYD	485	485	N-LINKED (CYCNA) (POTENTIAL).
F1	CARBOHYD	497	497	N-LINKED (CYCNA) (POTENTIAL).
NO	SYNDROME	618	65196	91FEAA60752B4F4B2.C6C64.

```

Query Match: 94.5% 4.88% Score 94.5% DB 1: Length 619;
Post Local Similarity: 25.28% Prod. No. 2.5;
Match es 60; Conservative 35; Moderate 24; Toxins 49; Gaps 14

QY 97 GGRPSM-----GQALYALAI PANF-----FFVATGRKQDPLVSLKMFLEKPAI----- 142
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 275 ECKAKIALI VNLGCI EIVASACGCTIVFVYKFDLELLTS-KISATIGMELLLEIF 333
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 141 -DTAAMGLAFTCLKSNSFNFGRRUKITMAIKTIVEELLKATIPGHHGVNYSPLALQF 201
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 330 EDIPEVPAI EICAYASCTI STASTS INAAAVTV-ETIKRPMPS-----IAPRTIV 385
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 207 LMTSPMGAEI--GTATKAPVALIASIDGAFQNALMSILE--PVLEKTYITLIFPD 257
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 486 I-----SRGTSPTGSAATIV-AALSILGGVIGDSFVYMYISRLIGAT-LEMLIPA 439
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 258 GIALPEVMIPEVAETIPGPAETISVTIGV---SLIPVYQGISVY---ASTIVEIVL 308
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 441 GNTGCVISGLAAG-----LAVSIWAVAGATIVPDPGPGVGHIPTSACCTIMSVL 489
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

DESCRIPT 14
DIDAT HOMON
ID: DIA_HOMON
AC: Q9NSV4
ID: 16-047-2001 (Ref. 40, Cited)
ID: 16-047-2001 (Ref. 40, last sequence update)
ID: 16-047-2001 (Ref. 40, last annotation update)
DB: Diphtherous Protein homolog 1 (G) aphamous related family 3 (DBPF3)

```

01. (Unannotated).
 02. DIAPH3 OR DIAP3.
 03. Homo sapiens (Human).
 04. Fukuyama; Molozay; Chordata; Crustacea; Vertebrata; Euteleostomi;
 05. Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 06. NCBI_TaxID:9606;
 07. [1]
 08. RE SEQUENCE FROM N.A. (ISOPFORM 2).
 09. TISSUE-Testis.
 10. Blum H., Buecherich S., Meeres H.-W., Gassschubert J., Wilmann S.;
 11. collected (Jan 2000) to the UniProt database.
 12. [2]
 13. RE SEQUENCE OF 468-853 FROM N.A. (ISOPFORMS 1 AND 2).
 14. Smith M.;
 15. Submitted (Jul-2000) to the EMBL/Genbank/DDB databases.
 16. -1- FUNCTION: KININS TO GTP-BIND FORM OF RHO AND TO FORMIN. ACTS IN
 17. A RHO-DEPENDENT MANNER TO RECRUIT PROTEIN TO THE MEMBRANE, WHERE
 18. IT PROMOTES ACTIN POLYMERIZATION. IT IS REGULATED FOR CYTOKINESIS,
 19. STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
 20. SERUM RESPONSE FACTOR. 148 PROTEIN CODING RHO AND SER- TYR-SINE
 21. KINASE JODING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS. (BY
 22. SIMILARITY).
 23. -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SEEN ON HERE) AND 2: ARE
 24. PRODUCED BY ALTERNATIVE SPLICING.
 25. -1- DDBM: DDBS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
 26. RHO-GTP ACTIVATES THE DDBS BY DISRUPTING THE GBD-DAD INTERACTION
 27. (BY SIMILARITY).
 28. -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
 29. -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 1 (FH1) DOMAIN.
 30. -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 2 (FH2) DOMAIN.
 31. -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLGY 4 (FH4) DOMAIN.
 32. -1- SIMILARITY: CONTAINS 1 DUF AUTOREGULATORY DOMAIN (DAD).
 33. -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLGY FAMILY, DIAPHANOUS
 34. SUBFAMILY.
 35. This SWISS-PROT entry is copyright. It is produced through a collaboration
 36. between the Swiss Institute of Bioinformatics and the EMBL outstation.
 37. the European Bioinformatics Institute. There are no restrictions on its
 38. use by non-profit institutions as long as its content is in no way
 39. modified and this statement is not removed, usage by and for commercial
 40. entities requires a license agreement. See: <http://www.ebi.ac.uk/Products/seqdata/doc/seqdata.html>
 41. or send an email to license@ebi.ac.uk.
 42. CC
 43. CC
 44. CC
 45. CC
 46. CC
 47. CC
 48. CC
 49. CC
 50. CC
 51. CC
 52. CC
 53. CC
 54. CC
 55. CC
 56. CC
 57. CC
 58. CC
 59. CC
 60. CC
 61. CC
 62. CC
 63. CC
 64. CC
 65. CC
 66. CC
 67. CC
 68. CC
 69. CC
 70. CC
 71. CC
 72. CC
 73. CC
 74. CC
 75. CC
 76. CC
 77. CC
 78. CC
 79. CC
 80. CC
 81. CC
 82. CC
 83. CC
 84. CC
 85. CC
 86. CC
 87. CC
 88. CC
 89. CC
 90. CC
 91. CC
 92. CC
 93. CC
 94. CC
 95. CC
 96. CC
 97. CC
 98. CC
 99. CC
 100. CC
 101. CC
 102. CC
 103. CC
 104. CC
 105. CC
 106. CC
 107. CC
 108. CC
 109. CC
 110. CC
 111. CC
 112. CC
 113. CC
 114. CC
 115. CC
 116. CC
 117. CC
 118. CC
 119. CC
 120. CC
 121. CC
 122. CC
 123. CC
 124. CC
 125. CC
 126. CC
 127. CC
 128. CC
 129. CC
 130. CC
 131. CC
 132. CC
 133. CC
 134. CC
 135. CC
 136. CC
 137. CC
 138. CC
 139. CC
 140. CC
 141. CC
 142. CC
 143. CC
 144. CC
 145. CC
 146. CC
 147. CC
 148. CC
 149. CC
 150. CC
 151. CC
 152. CC
 153. CC
 154. CC
 155. CC
 156. CC
 157. CC
 158. CC
 159. CC
 160. CC
 161. CC
 162. CC
 163. CC
 164. CC
 165. CC
 166. CC
 167. CC
 168. CC
 169. CC
 170. CC
 171. CC
 172. CC
 173. CC
 174. CC
 175. CC
 176. CC
 177. CC
 178. CC
 179. CC
 180. CC
 181. CC
 182. CC
 183. CC
 184. CC
 185. CC
 186. CC
 187. CC
 188. CC
 189. CC
 190. CC
 191. CC
 192. CC
 193. CC
 194. CC
 195. CC
 196. CC
 197. CC
 198. CC
 199. CC
 200. CC
 201. CC
 202. CC
 203. CC
 204. CC
 205. CC
 206. CC
 207. CC
 208. CC
 209. CC
 210. CC
 211. CC
 212. CC
 213. CC
 214. CC
 215. CC
 216. CC
 217. CC
 218. CC
 219. CC
 220. CC
 221. CC
 222. CC
 223. CC
 224. CC
 225. CC
 226. CC
 227. CC
 228. CC
 229. CC
 230. CC
 231. CC
 232. CC
 233. CC
 234. CC
 235. CC
 236. CC
 237. CC
 238. CC
 239. CC
 240. CC
 241. CC
 242. CC
 243. CC
 244. CC
 245. CC
 246. CC
 247. CC
 248. CC
 249. CC
 250. CC
 251. CC
 252. CC
 253. CC
 254. CC
 255. CC
 256. CC
 257. CC
 258. CC
 259. CC
 260. CC
 261. CC
 262. CC
 263. CC
 264. CC
 265. CC
 266. CC
 267. CC
 268. CC
 269. CC
 270. CC
 271. CC
 272. CC
 273. CC
 274. CC
 275. CC
 276. CC
 277. CC
 278. CC
 279. CC
 280. CC
 281. CC
 282. CC
 283. CC
 284. CC
 285. CC
 286. CC
 287. CC
 288. CC
 289. CC
 290. CC
 291. CC
 292. CC
 293. CC
 294. CC
 295. CC
 296. CC
 297. CC
 298. CC
 299. CC
 300. CC
 301. CC
 302. CC
 303. CC
 304. CC
 305. CC
 306. CC
 307. CC
 308. CC
 309. CC
 310. CC
 311. CC
 312. CC
 313. CC
 314. CC
 315. CC
 316. CC
 317. CC
 318. CC
 319. CC
 320. CC
 321. CC
 322. CC
 323. CC
 324. CC
 325. CC
 326. CC
 327. CC
 328. CC
 329. CC
 330. CC
 331. CC
 332. CC
 333. CC
 334. CC
 335. CC
 336. CC
 337. CC
 338. CC
 339. CC
 340. CC
 341. CC
 342. CC
 343. CC
 344. CC
 345. CC
 346. CC
 347. CC
 348. CC
 349. CC
 350. CC

Query Match	4.8%	Score 93	Pos 1	Length 853	
Best Local Similarity	23.6%	Pred. No. 4.3			
Matches	53	Conservative	45	Mismatches	103
				Indels	40
				Gaps	92
179 MATPVEFETKAGTPEHCVNVPVLAIGTMTSDMRCACTGACGKAVVALASDGD	279				

[illegible]

• ,

36 STRAIN 0706/957. PubMed-11139578.
 37 MEDLINE 21179132.
 38 YAN Q., LIOU H., LEE J.-D., ABE J., BECK B.C.
 39 "Molecular Cloning of Mouse EEF2/SMB1 Effector Variant 5 and
 40 Characterization of EEF2 Functional Domains." *J. Biol. Chem.* 276:10487-10498(2001).
 41
 42 1. *FUNCTION: AKA4-455.1* IS A MEMBER OF THE FAMILY OF PROTEIN KINASES.
 43
 44 *ENZYME: AKA4-455.1* EC 2.7.11.1.
 45 *ISSN: 01657939*. IMFC.
 46
 47 *NCBI MOTIF 140447* MAPK7.
 48
 49 *LOCATED: 1* PROTEIN: 1, PK- γ -kinase.
 50
 51 *LOCATED: 1* PROTEIN: 3227, MAP-kin.
 52
 53 *LOCATED: 1* PROTEIN: 2965, P-38 α extension.
 54
 55 *LOCATED: 1* PROTEIN: 3716, RNA pol omega.
 56
 57 *LOCATED: 1* PROTEIN: 2579, Raf-1 kinase.
 58
 59 *LOCATED: 1* PROTEIN: 669, p44mapk.
 60
 61 *LOCATED: 1* PROTEIN: 217, PRK1DEXTIN3.
 62
 63 *SMART: SM00220*. S_AKc_1.
 64
 65 *PROSITE: PS01451*. MAPK. 000749.1.
 66
 67 *PROSITE: PS00011*. PROTEIN KINASE. 0001.1.
 68
 69 *PROSITE: PS00108*. PROTEIN KINASE. ST. 1.
 70
 71 *MAP Kinase: 1* kinase. serine/threonine tyrosine kinase. Transferase.

ATP-binding, kinase, serine/threonine protein kinase; transferase.

DR PROSITTE; PS001108; PR001109


```

Query Match: 4.88; Score 93.5; DB 11; Length 806;
Host Local Similarity 19.28; Pred. No. 20;
Matches 84; Conservative 51; Mismatches 128; Indels 171; Gaps 19;

QY 22 ETENSHLVKKKGGHLEMMRSELEHNSIYVCHLSLQAGIKHGYHSLKIQYQ 81
DB 138 DIMSHLMQIHSQ-----P-LLEHRYFLYQILK-----GLK---YHSAQVLR 181
QY 82 GQTHGSAFSETEGGGKPSMQD-ALYLLAPARCPVPCWKG-----LELVS 129
DB 182 DQ-----KERR-----LLVNRCHTKLEPCMAHLCCTSTAEHREYET 219
QY 130 Q-----LKNP-----LEKKKALDIAAMGLAFTLKPSNENPCPR-----QRLIMAIR 173
DB 220 FYVAFQWVPALPELMSEHVEYQALPMSVQTHQPMARPGQLEPKRYVHIGLIMVGLG 279
QY 174 TVPEETLKAGTFECHPCNVYGP-----LAKGTGRTIMGALLQTA 215
DB 280 TPSAVVIGAVAFVAVATQSTPPQGPVPEYVPCANIGQALSLSPMLPEFSAPISAA 339
QY 216 QL-----KARVALLASLD-----SAFQVA 235
DB 340 ALIRHPEIAYRHRRGRRGAPPPGVAFTDFALIGPTKEFAVAPPPHAPPEPI 399
QY 236 LMTSGLLPVNHRKTYTHLP-----DGL--APRVMLHMAETTYVQRIISVTLQ-- 284
DB 400 KQPSLQPVVSEPCPVHFMSPMPSPGDCAMESPAPALPPGSDPAPPT--VDLTLQPA 456
QY 295 VSTLPEVQSTSVAGTVELFAHHLDPPEVQVASTSDGLSWAGAGERE 342
DB 457 PPAHLARFKE-----GAINSHKAAALAAALNTLSGP----- 490
QY 443 FVGLLRQINTEPL 354
DB 491 ----LRQPSAPL 499

```

Search completed: September 4, 2002, 11:48:26
 Job time: 7624 sec



lung cancer associated gene sequences, referred to as lung cancer
 antigens, useful for treatment, prevention, and diagnosis of disorders
 such as lung cancer.

claim 11: Page 954-955; 1425pp; English.

Polynucleotide sequences AAF1942 - AAF1424 encode human lung cancer
 associated proteins represented in AAF58106 - AAF58548. Lung cancer
 associated proteins and polynucleotide sequences, their agonists, and
 antagonists may have neuroprotective, cytoprotective, cardioprotective,
 immunomodulatory, muscular active, general, vitelline, gastrointestinal
 activity, nephroprotective and infective, gynecologically or antibacterial
 activity. The invention also includes antibodies specific for the
 protein or polynucleotide sequences. The lung cancer associated
 polynucleotide sequences may be used for detection of lung cancer,
 chromosome identification, as chromosome markers, and for numerous other
 diagnostic or research purposes. The proteins may be used to treat
 disorders such as neural, immune, muscular, reproductive,
 gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
 disorders. The proteins may also be used in the treatment of wounds and
 infectious diseases. Polynucleotide sequences AAF1425 - AAF1433 and
 peptide AAF58549 are used in the course of the invention for the
 identification and characterization of the polynucleotide and protein
 sequences.

Sequence 268 AA:

Query Match 41.6%; Score 803.5; DR 21; Length 268;
 best local similarity 74.5%; Prod No. 7-96-78;
 Matches 161; Conservative 0; Mismatches 4; Indels 51; Gaps 1;

QY 1 MRHIAFLFLGVALTEMCETPEMDSHVEKLGQHLPMMDRLSLEHNPSTVIGRL 60
 |||||
 Db 39 mrlhgaflflgvaltelmcelpemshlvexlqhlipmdrlstehnpstlygdl 98
 |||||
 QY 61 SSLDAGTKPKYIHSTKGYGCGTIGSAPSEDCGTVGGRSMGQALLYLALPANCEFVR 120
 |||||
 Db 99 ssldagtkedyihstlkgygclgsatseddqdkqkpsmqxllalrancelvx 158
 |||||
 QY 121 GHRKPIVLSQKWELEDEKPAI 142
 |||||
 Db 159 ghrkplvlsqkweledeksygytqtlldlhlphkrlvdsyvdsl 216
 |||||
 QY 143 -----DTAAMAGIAFTCLKRSNPNQKR 165
 |||||
 Db 219 yavepibqhbhsyvdtaamagiatclkrstnpqr 254
 |||||

RESULT 2

AA300285 standard; Protein: 54 AA.

AA300285;

06-CT-2000 (first entry)

Human secreted protein, SEQ ID NO. 4366.

Human, 5' EST, expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping.

Human sapiens.

EP1634401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0206610.

26-FEB-1999; 99US-0122487.

(BEST) GENSET.

Human Milne Edwards J. Inherent A. Giordano J;
 WPI: 2000-500481/45.
 N-PSDB: AAC00291.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 diagnostic, forensic, gene therapy and chromosome mapping procedures

Claim 13: SEQ ID 4366; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number
 of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 were prepared from total human RNAs or poly(A) RNAs derived from 30
 different tissues. EST sequences usually correspond mainly to the 3'
 untranslated region (UTR) of the mRNA because they are often obtained
 from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 those cases where longer cDNA sequences have been obtained, the full 5'
 UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 ends and can therefore be used to obtain full length cDNAs and genomic
 DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 chromosome mapping procedures. They are used to obtain upstream
 regulatory sequences and to design expression and secretion vectors.

Sequence 54 AA:

Query Match 14.8%; Score 285; DR 21; Length 54;
 best local similarity 100.0%; Prod. No. 5-20-24;
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRHGAFLFLGVALTEMCETPEMDSHVEKLGQHLPMMDRLSLEHNPSTVIGRL 54
 |||||
 Db 1 mrlhgaflflgvaltelmcelpemshlvexlqhlipmdrlstehnpst 54
 |||||

RESULT 3

AA32711 standard; Protein: 399 AA.

AA32711;

17-JUN-1993 (first entry)

Human intrinsic factor.

Human, intrinsic factor; PCR primer; polymerase chain reaction;

M13mp19; Phage vector; Stomach; cDNA library; probe; Vitamin B12.

Human sapiens.

JP05015375-A.

26-JAN-1993.

11-JUL-1991; 91JP-0171062.

11-JUL-1991; 91JP-0171062.

(KANF) KANEKA CORP.

WPI: 1993-070182/09.

P-PSDB: AAR32711.

DNA sequence coding human intrinsic factor used for

quantitative analysis of vitamin-B12

Claim 1: Page 4-6; 10pp; Japanese.

This sequence represents human intrinsic factor. The coding sequence
 was isolated by PCR using the primer sequences given in AAG37410-11.

CC The amplified DNA was subcloned into M13mp19 phage vector and the
 CC DNA sequence determined. This sequence was cloned from a human
 CC stomach cDNA library using the probes given in AAK37312-13. Human
 CC intrinsic factor protein can be used for quantitative analysis of
 CC vitamin B12.

XX Sequence 399 AA;

Query Match 14.78; Score 265; DR 14; Length 399;
 Best Local Similarity 25.29; Pred. No. 1,79-19;
 Matches 104; Conservative 61; Mismatches 149; Indels 100; Gaps 15;

```

UY 21 CEIPMSHSLEVKIKQHLIMMDRLSLEHNPSTVGLSSDQATKRLDELHSLKAY 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 CSYSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 81 QWPTTCSAFSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 54 GKLTLYGLMSDNDI 1191911000111111111111111111111111111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 138 -----EKRA-----UTAMAGIAPIC 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 105 ASPMAEASATYPSLALALQKQSEATPIAVTAKLLAUSPSIYVDIQAALALIC 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 155 LKES NENPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 165 MYKLTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 212 TATATATATATATATATATATATATATATATATATATATATATATATAT
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 KQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 271 --1111111111111111111111111111111111111111111111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 272 QH1111111111111111111111111111111111111111111111
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 316 G-FTYEQASISQPLYSV--MGKAGEBEPWOLLRDPTPLLOGLAVYPRDGETI 369
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 337 PWTIKTATMSAGTSSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 4

AAK32712 standard; Protein: 421 AA

AC AAK32712;

DI 17-JUN-1993 (first entry)

DE Rat intrinsic factor.

KM Rat intrinsic factor PCR primer; polymerase chain reaction;

KM Multiple phage vector; stomach cDNA library; phage; vitamin B12.

OS Rattus norvegicus.

PN JP05015475-A.

PD 26-JAN-1993.

PE 11-JUL-1991; 91JP-0171062.

PR 11-JUL-1991; 91JP-0171062.

PA (KANF) KANEKA CORP.

DB WPI - 1992-070182,009

DB P-PSDB; AAK32711.

PT DNA sequence coding human intrinsic factor - used for

PI quantitative analysis of vitamin-B12

PS Disclosure; Page 7-8; 10pp; Japanese.

XX This sequence represents rat intrinsic factor. The sequence was
 CC determined using the human intrinsic factor sequence. The human
 CC sequence was isolated by PCR using the primer sequences given in
 CC AAK37310-11. The amplified DNA was subcloned into M13mp19 phage
 CC vector and the DNA sequence determined. The full length sequence was
 CC cloned from a human stomach cDNA library using the probes given in
 CC AAK37312-13. Human intrinsic factor protein can be used for quantitative
 CC analysis of vitamin B12.

XX Sequence 421 AA;

Query Match 13.58; Score 261.5; DR 14; Length 421;
 Best Local Similarity 23.76; Pred. No. 4,49-19;
 Matches 104; Conservative 71; Mismatches 152; Indels 111; Gaps 17;

```

UY 9 FTLAVCANP-----PWRPFRPMSGNVFKICQGNTPWQWQSG-----PPIAP 52
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 YLLVNVKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 53 STVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 SLLIAMLAS-----LYLQAKKLTLYELMSDSDI-----LUGLALALAL 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 113 RANCFYVQKIKRELVSQK-W----- 133
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 TSSQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 134 ---FTRPKRATPAMAGAPIC-----KPSNENPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 KLLMMSSTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 184 TPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 219 KQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 244 VLNRKTYTDL-----TPDCLAPRWML-----PAETTPQVQVQVQVQV
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 SLKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 295 STSVIAGSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 338 QKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
UY 352 TPLVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 397 TPLQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 5

AAK37733 standard; Protein: 421 AA.

AC AAK37733;

DI 23-JUL-1993 (first entry)

DE Rat intrinsic factor.

KM Human intrinsic factor; phage clone; rat.

OS Rattus norvegicus.

PN JP05049478-A.

PD Key

DB Key

DB Key

DB Key

DB Key

15	22	FEF	2000	2600M	US04414	
16	24	FEF	2000	2600M	US04414	
17	24	FEF	2000	2600M	US05004	
18	01	MAR	2000	2000M	US05001	
19	20	MAR	2000	2000M	US05001	
20	21	MAR	2000	2000M	US05001	
21	21	MAR	2000	2000M	US05001	
22	17	MAY	2000	2000M	US13705	
23	22	MAY	2000	2000M	US13705	
24	30	MAY	2000	2000M	US14941	
25	02	SEP	2000	2000M	US15264	
26	10	SEP	2000	2000M	US15264	
27	10	SEP	2000	2000M	US15264	
28	10	SEP	2000	2000M	US15264	
29	10	SEP	2000	2000M	US15264	
30	10	SEP	2000	2000M	US15264	
31	10	SEP	2000	2000M	US15264	
32	10	SEP	2000	2000M	US15264	
33	10	SEP	2000	2000M	US15264	
34	10	SEP	2000	2000M	US15264	
35	10	SEP	2000	2000M	US15264	
36	10	SEP	2000	2000M	US15264	
37	10	SEP	2000	2000M	US15264	
38	10	SEP	2000	2000M	US15264	
39	10	SEP	2000	2000M	US15264	
40	10	SEP	2000	2000M	US15264	
41	10	SEP	2000	2000M	US15264	
42	10	SEP	2000	2000M	US15264	
43	10	SEP	2000	2000M	US15264	
44	10	SEP	2000	2000M	US15264	
45	10	SEP	2000	2000M	US15264	
46	10	SEP	2000	2000M	US15264	
47	10	SEP	2000	2000M	US15264	
48	10	SEP	2000	2000M	US15264	
49	10	SEP	2000	2000M	US15264	
50	10	SEP	2000	2000M	US15264	
51	10	SEP	2000	2000M	US15264	
52	10	SEP	2000	2000M	US15264	
53	10	SEP	2000	2000M	US15264	
54	10	SEP	2000	2000M	US15264	
55	10	SEP	2000	2000M	US15264	
56	10	SEP	2000	2000M	US15264	
57	10	SEP	2000	2000M	US15264	
58	10	SEP	2000	2000M	US15264	
59	10	SEP	2000	2000M	US15264	
60	10	SEP	2000	2000M	US15264	
61	10	SEP	2000	2000M	US15264	
62	10	SEP	2000	2000M	US15264	
63	10	SEP	2000	2000M	US15264	
64	10	SEP	2000	2000M	US15264	
65	10	SEP	2000	2000M	US15264	
66	10	SEP	2000	2000M	US15264	
67	10	SEP	2000	2000M	US15264	
68	10	SEP	2000	2000M	US15264	
69	10	SEP	2000	2000M	US15264	
70	10	SEP	2000	2000M	US15264	
71	10	SEP	2000	2000M	US15264	
72	10	SEP	2000	2000M	US15264	
73	10	SEP	2000	2000M	US15264	
74	10	SEP	2000	2000M	US15264	
75	10	SEP	2000	2000M	US15264	
76	10	SEP	2000	2000M	US15264	
77	10	SEP	2000	2000M	US15264	
78	10	SEP	2000	2000M	US15264	
79	10	SEP	2000	2000M	US15264	
80	10	SEP	2000	2000M	US15264	
81	10	SEP	2000	2000M	US15264	
82	10	SEP	2000	2000M	US15264	
83	10	SEP	2000	2000M	US15264	
84	10	SEP	2000	2000M	US15264	
85	10	SEP	2000	2000M	US15264	
86	10	SEP	2000	2000M	US15264	
87	10	SEP	2000	2000M	US15264	
88	10	SEP	2000	2000M	US15264	
89	10	SEP	2000	2000M	US15264	
90	10	SEP	2000	2000M	US15264	
91	10	SEP	2000	2000M	US15264	
92	10	SEP	2000	2000M	US15264	
93	10	SEP	2000	2000M	US15264	
94	10	SEP	2000	2000M	US15264	
95	10	SEP	2000	2000M	US15264	
96	10	SEP	2000	2000M	US15264	
97	10	SEP	2000	2000M	US15264	
98	10	SEP	2000	2000M	US15264	
99	10	SEP	2000	2000M	US15264	
100	10	SEP	2000	2000M	US15264	

Query Match 4.88; Score 93.5; DB 22; Length 511;
Best Local Similarity 18.5%; Prod. No. 0.77;
Matches 77; Conservative 51; Mismatches 125; Indels 163; Gaps 20;

23 IPHNDSHLVKKGQMLIPWMDKSLPHNLSYVGLRSLSSQACSTKKNVTHLKLQYGG 82
11 11

0Y 258 CLAPVMEFAETIPQIGELISVHOV---LSLLPYROSTISVL---AGSTVEDVL 308
| | | | | : | | | | | : | | | | | : | | | | |
Dh 440 CMTPTISQIAG--- larslwagvqatlppqetmg;lpsdagstlactl 189

Search completed: September 4, 2002, 09:42:41
Job time: 6028 sec



RESULT 6
 H72031
 Lipoteic acid synthetase (lipA) (inferred) Chlamydia pneumoniae (Strains: W629, 363)
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999, seq:seqs-protein-23 Apr 1999, text_change:1 May 2000
 C:Accession: H72031, C81509
 R:Kakman, S.; Mitchell, W.; Marchhe, R.; Lamuel, C.; Fan, J.; Ollinger, L.; Greenwood, J.
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A720001, MIMD:20150255
 A:Accession: H72031
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <AAK>
 A:Cross-references: GH:AE001663, GH:AE001663, NID:0437134, PIDN:AA018969.1, PID:043771
 A:Experimental source: strain: W629
 A:Method: L.P.; Brinkham, R.C.; Shen, C.; Gill, S.R.; Holscher, J.P.; Kistner, J.; Hickey,
 C.; Dodson, E.; Quinn, M.; Nelson, W.; Deery, R.; Kolonay, J.; McCarthy, G.; Marchhe,
 R.; Nature Genet. Res. 28, 1397-1406, 2000
 A:Title: Genome sequence of Chlamydia pneumoniae W629 and identification of genes
 A:Reference number: A815003, MIMD:20150255
 A:Accession: C81509
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <KFA>
 A:Cross-references: GH:AE000913, GH:AE000913, NID:0189450, PIDN:AA139833.1, PID:018939
 A:Experimental source: strain: AR39, HU cells
 C:Genetics:
 A:Genes:
 A:Support family: Lipoteic acid synthetase

[illegible]

DB 541 HWATKISQHEVEV:DOJFO... NLEKPF TSECKNLSIKVYDWLIRPLETO 589

OY 454 LLO 356

DB 590 LSO 592

Search completed: September 4, 2002, 09:41:19
Job time: 6851 sec

[illegible]

OY	1468	caaaacacgaatccgcacagcgcgcgacgatcttctctgttttgqqt	1527
Db	1628	CAAAAAAGGACGTGGCGAAGCCGCAGTGTGTGAAGAATCATCTTCCTCCTGCTTGAGGAT	1687
OY	1528	ccttcacaaaaggactctctcagcgcggcgagcatgaccttaccagctctcactctcgc	1587
Db	1698	CCTTCACAAGAAAGGCCCTGCTCAGGCGCGGAGGCTATGAGGCGCTCAGGCGCACCTTCCTACTCTGCG	1747
OY	1568	tatttagatctaacacctctcagctgatttatgagcaatactaatctgagaactctacgtgagac	1647
Db	1748	TGTTAGAGTGCACAGCTCTTAGGCTGCTCTGGCAACAGAGTGGAGCAACATCTGAGTGAACAT	1807
OY	1648	tactcatctactgctctcagacaatatctaaagatctgaatgacacctgtaaaaaaataaa	1706
Db	1858	TGCTCAGCTGCTTGCTCTCAGCAAAAATTAAAGATTATGATGGCTGTGTGAGTCTGTGTAAMAAA	1856
RESULT	5		
HOMOTCI			
LOCUS		1866 bp mRNA linear 14-JAN-1995	
DEFINITION		Human transcobalamin II (TCII) mRNA, complete cds.	
ACCESSION		M00396	
VERSION		M00396.1 GI:339195	
KEYWORDS		transcobalamin II,	
SOURCE		human umbilical vein endothelial cell; cDNA to mRNA.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
		1 (theses 1 to 1866)	
		Platica,O., Janerzko,K., Quadros,E.V., Kopec,A., Romalin,K. and	
		Kotchenberg,S.P.	
		The cDNA sequence and the deduced amino acid sequence of human	
JOURNAL		transcobalamin II show homology with rat nutritive factor and human	
MEETING		transcobalamin I	
FEATURES		J Biol Chem 266 (17), 7860-7863 (1991)	
Source		Location/Qualifiers	
		1..1866	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/map_229"	
		/cell_type="endothelial cell"	
		/tissue_type="umbilical vein"	
		1..1866	
		/gene="TCN2"	
		38..1421	
		/gene "TCN2"	
		/codon_start 1	
		/product="transcobalamin II"	
		/protein_id="AA01054.1"	
		/db_xref="GI:339195"	
		38..91	
		38..91	
		/gene="TCN2"	
		/note="G00-119 608"	
		92..1318	
		/gene="TCN2"	
		/note="G00-119-608"	
		/evidence=experimental	
		/product="transcobalamin II"	
BASE COUNT		395 a 575 c 500 g 396 t	
ORIGIN			

[illegible][illegible]

[illegible][illegible]

Template: 10 ng
 Primer: each 5 µM
 dNTPs: each 4 mM
 Tag Polymerase: 0.025 units/ul
 Total Vol: 20 ul

Buffer:

MgCl₂: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 9.3

FEATURES

SOURCE

location/qualifiers
 1..349
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="768_(B,C)_7"

SIZE

primer_bind 67..86
 primer_bind 67..319

primer_bind complement(300..319)

HASTE_COINP 68 a 110 c 101 g 70 t

ORIGIN

Query Match

20.2%; Score 347.4; DB 11; Length 349;

Best Local Similarity 99.7%; Pred. No. 5.4e-79;

Matches 348; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

UY 1232 gattaccccaagcccttcgaaggaacctataccatgacacacttgagagcagaga 1351
DB 1 GATACCCCAAGCCCTTCGAAGGACCTATACCAATGACCACTTGAGAGCAGAGA 60
UY 1352 gccaaagatcttcctggaagatcttctcggccaagctcgccagctgacaggt 1411
DB 51 GCCAAGCATCTTCCTGGGAAGTCTTGGGCAATCTGACCAATCTGAGAGGAG 120
UY 1412 ctccgaatgaagagcccccctgctcgaatgagcagatgaagacatcctctgcaaa 1471
DB 121 CTCCGATGAAGAGCCCCCTAGTCTGATGAGCAATGAGCAATCTGAGAGGAG 180
UY 1472 aaacgaatccgaagccgaagatgttgaagaaccacactcattctatggttgaggtccg 1531
DB 131 AAACGAGTCCGACGACGCGCAGATTTGGAATGACATCTGTTGCTTGAAGTCTTG 249
UY 1532 caaagaagctctcctgaagccggaggtatagacctaaaccagctctccactcgtgtt 1591
DB 241 CAAGAAGCTCTCCTGAAGCCGGAGGTATAGACCTAAACCAAGCTCTCCACTCGTGT 400
UY 1592 aagatggaagctccgaagctgttctggaacagtagctggggaagacctca 1640
DB 301 AAGATGGAAGCTCCGAAGCTGTCTTGTGAGTACCTGGGGAAGACCTCA 349

```

Search completed: September 4, 2002, 06:35:49
 Job time: 10675 sec

[illegible]

[illegible]

XX	AAAF68505/c	5	RESULT	5	
XX	ID	AAAF68505	standard; cDNA: 338 BP.		
XX	AC	AAAF68505;			
XX	XX	12-Apr-2001	(first entry)		
XX	XX	Human lung tumour protein related nucleotide sequence SPO ID NO:440.			
XX	XX	Human lung cancer; lung tumour; lung tumour protein; gene therapy; lung cancer antigen; lung tumour specific antigen; diagnosis; vaccine; cytotoxic; antitumor; inhibition; ss.			
XX	XX	Human sapiens.			
XX	XX	W0200100928-A2.			
XX	XX	04 JAN 2001.			
XX	XX	30-JUN-2000; 2000WO-00518061.			
XX	XX	30-JUN-1999; 990S-0346492.			
XX	XX	15-OCT-1999; 990S-0419456.			
XX	XX	17-DEC-1999; 990S-0466867.			
XX	XX	30-DEC-1999; 990S-0476400.			
XX	XX	04-MAR-2000; 2000US-0519642.			
XX	XX	22-MAR-2000; 2000US-0534077.			
XX	XX	10-APR-2000; 2000US-0546259.			
XX	XX	27-APR-2000; 2000US-0550406.			
XX	XX	06-JUN-2000; 2000US-0589184.			
XX	XX	(CORI-) CORIXA CORP.			
XX	XX	Wang T, Banour GS, Lodes MJ, Fanger GK, Vedvick TS, Carter D;			
XX	XX	Keller MW, Mannion J;			
XX	XX	WPI: 2001-071488/08.			
XX	XX	lung tumor-associated proteins and the nucleic acids that encode them,			
XX	XX	used for preventing, diagnosing and treating lung cancer.			
XX	XX	claim 4: Page 300-401; 43pp; English.			
XX	XX	The present invention describes immunogenic portions of lung tumour-			
XX	XX	associated proteins (1) and the nucleic acids (NAs) that encode them.			
XX	XX	(1) have cytotoxic activity and can be used in gene therapy, antisense			
XX	XX	inhibition and in vaccines. The NAs and the lung tumour associated			
XX	XX	proteins they encode may be used in the prevention, treatment and			
XX	XX	diagnosis of diseases associated with their inappropriate expression,			
XX	XX	especially lung cancers. For example, the NAs may be administered to			
XX	XX	treat diseases by rectifying mutations or deletions in a patient's genome			
XX	XX	that affect the activity of the protein by expressing inactive proteins			
XX	XX	or to supplement the patient's own production of (1). Additionally, the			
XX	XX	NAs may be used to produce the lung-tumour associated protein, according			
XX	XX	to standard recombinant DNA methodology. Conversely, antisense NA			
XX	XX	molecules may be administered to down regulate protein expression by			
XX	XX	binding with the cells own genes and preventing their expression. The NA			
XX	XX	and complementary sequences may also be used as RNA probes in diagnostic			
XX	XX	assays to detect and quantify the presence of similar NA sequences in			
XX	XX	samples, and hence which patients may be in need of treatment for lung			
XX	XX	cancer. The (1) may be used as antigens in the production of antibodies			
XX	XX	and in assays to identify modulators (agonists and antagonists) of the			
XX	XX	expression and activity of the protein. AAF68084 to AAF68878 and			
XX	XX	AAF68648 to AAF68678 represent human lung tumour protein related			
XX	XX	nucleotide and protein sequences which are used in the exemplification			
XX	XX	of the present invention.			
XX	XX	Sequence 348 BP; 68 A; 81 C; 117 G; 71 T; 1 other;			
XX	XX	Query Match	19.6%	Score 347	len 22
XX	XX	Best Local Similarity	99.78%	Prod No. 1,46-75;	Length 438;

	Matches	437	Conservative	9	Mismatches	1	Indels	0	Gaps	0
CY	1045	gagatgaattgggttttcttgacattcttcgaagtcccaatcccgcctggttggaattact	1104							
DB	348	GGAGAAAGGAGATCTCGACGCTTCTCCGAGGCCAACCACCACTGTTCGAAGATTAT	279							
CY	1105	actgaatcctgatcctctcatgataatgaaatcttgaatatctgaatcttgaatcttgaat	1164							
DB	278	GTGACTATCATCCTCCATGCATGATGAAAACCAATTAAGCTTAGCTTGTTATGCTGACCT	219							
CY	1165	cctgacct	1224							
DB	218	CTGAGCTTCCTCCCTTCCTCCAGCACCTTCGACACCTCCCTGAGGCTTCCTAACCTCCCTCTGAT	159							
CY	1225	gtccccctgagctggaggagctctgacacacacacacacacacacacacacacacacacacac	1284							
DB	158	GTCCTCTGCACATCAGTAATCTGCCTCTGACCTCTCTGACCTCTCTGCTGCTGCTTGAACAATGC	99							
CY	1285	cctccctgaatcacccac	1344							
DB	98	CCCCCTTATATCATCTCCATGACATGACATGACATGACATGACATGACATGACATGACATGAC	39							
CY	1345	gcagaagaccgaagatctctcccttagaagaatctctctga	1382							
DB	48	GTACACAGATTAAGTATCTTCT	1							
RESULT	6									
ID	AABH7310									
XX	AABH7310 standard; DNA; 344 bp									
XX	AABH7310:									
XX	27-JUN-2002 (first entry)									
DE	Human single nucleotide polymorphism containing DNA sequence #2167.									
RW	Predicted marker: polymorphisms found disease diagnosis: Trexler's									
KM	phenotypic trait: gene therapy: for genetic pathology: mapping: cancer;									
FW	response: on the molecularly polymorphic: SNP, db.									
XX	Homo sapiens.									
XX	Key									
EH	Variation									
F1	Location/units									
F1	replace(172:A)									
F1	/seq_a									
XX	standard_name= "single nucleotide polymorphism"									
XX	WG96 contig A2									
XX	21-oct-1999.									
XX	30-MAR-1999.									
XX	09-APR-1998.									
XX	(WHIP) WHITEHEAD INST BIOMEDICAL RES.									
XX	Lander ES, Wang D, Hudson J.									
XX	WPL 1999 (2017/7).									
XX	Polymorphic human genomic sequences and related allele-specific probes									
XX	and primers, useful for genetic analysis, e.g. diagnosis and monitoring									
XX	of disease									
XX	Claim 1: Page 271: 39pp; English.									
XX	This invention describes novel human nucleic acid segments (1)									
XX	containing polymorphic sites. The polynucleotides of (1) are used for,									
XX	e.g. correlating disease polymorphisms (or disease susceptibility) or									
XX	other phenotypic traits (e.g. baldness, obesity, fertility, sterility,									
XX	response to drugs etc.); diagnosis and monitoring e.g. cancer,									

CC	inflammation, heart or central nervous system diseases; detecting
CC	susceptibility to microbial infection; treating or preventing such
CC	diseases; forensic analysis; gene therapy; paternity testing; mapping
CC	genomic loci associated with phenotypic traits (and subsequent cloning
CC	of the genes responsible); and the production of transgenic organisms.
CC	Antibodies raised against (1) are useful as diagnostic and therapeutic
CC	tools and in drug screening. AABH7616 : AAH87644 represent the human
CC	DNA sequences containing biallelic polymorphic sites described in the
CC	invention.
XX	
XX	Sequence 344 bp; 65 A; 106 C; 95 G; 68 T; 0 other;
XX	
XX	Query Match 19.3% Score 332.4; DH 20; length 344;
XX	Best Local Similarity: 39.7%, Evid. No. 1.9e-74;
XX	Matches 333; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
YY	1202 gattacagcaggccgacagggccctcgagagtcgcatacgaatggcccaccttgagacagaga 1351
YY	
YY	1 gatcaaccgagccacaagactcttcttagajjccctataacatbajjccaaatggagacagaga 60
YY	1352 gccacagatctctccagggaactctttctgccaagtctcacacagctgagccctgcagat 1411
YY	
YY	61 gccacagatctctctcagggaactctttctgccaagtctcacacagctgagccctgcagat 120
YY	1412 cagccat taagggagccgcac tttatctgatgtcatgaacgtctctgagagcccttggaaa 1471
YY	121 ctccatgaagcgcacccatgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 180
YY	1472 aaacagatctctcagacacatcttctt gaadacatctctctcgtcgtcgtcgtcgtcgtcgt 1531
YY	
YY	181 aaacagatctctcagacacacatcttctt gaadacatctctctcgtcgtcgtcgtcgtcgtcgt 240
YY	1532 caaacagatctctctcagacacacatcttctt gaadacatctctctcgtcgtcgtcgtcgtcgtcgt 1591
YY	241 caaacagatctctctcagacacacatcttctt gaadacatctctctcgtcgtcgtcgtcgtcgtcgt 600
YY	1592 gacagatctctcagacacacatcttctt gaadacatctctctcgtcgtcgtcgtcgtcgtcgt 1655
YY	
YY	342 agagtaaatctctcagacacacatcttctt gaadacatctctctcgtcgtcgtcgtcgtcgtcgt 344
YY	
RESULT 7	
AAH87616	
ID	AAH87616 standard; DNA; 344 bp.
AC	
AAH87616:	
CC	
CT	27-FEB-2002 (first entry)
II	
XX	
DE	Human single nucleotide polymorphism containing RNA sequence #2473.
XX	
PW	Biallelic marker; polymorphism; human disease diagnosis; treatment;
PW	phenotypic trait; gene therapy; forensic; paternity mapping; cancer;
KW	transient; single nucleotide polymorphism; SNP; ss.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
PH	replac(184..)
FP	Variation
TI	/tag
FT	/standard_name="single nucleotide polymorphism"
XX	
EN	W0995.095.A2.
XX	
PD	21-OCT-1999.
XX	
PF	30-MAR-1999; 99W-US06893.
XX	
PK	09-APR-1998; 98US-0057871.
XX	
FA	(WHED) WHITEHEAD INST BIOMEDICAL RES.
XX	

Wed Sep 4 13:30:57 2002

us-09-841-158-1.rng

Page 13

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 587)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gaisel, G., Jost, S.,
 Kucuk, T., Lacy, M., Le, N., Lemons, D., Maria, M., Martin, J., Moore, R.,
 Schollhorn, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wyllie,
 J., Worreston, P., and Wilson, P.
 TITLE Wash-Merck EST Project 1997
 JOURNAL Capitalized (1997)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty free through LINC; contact the
 IMAGe Consortium (info@image.llnl.gov) for further information.
 Seq primer: 41m13 fwd. ET from Amersham
 High quality sequence stop: 381.
 FEATURES
 SOURCE
 1..587
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="Soares_NSHMP_9w_01_PA_P_S1"
 /issue_type="Pooled human melanocyte, fetal heart, and
 /lab_host="PH10B"
 /note="Organ: mixed (see below): Vector: pTZ19-Pac
 (Pharmacia) with a modified polylinker; Site_1: Not 1;
 Site_2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NbH, pregnant uterus
 NHMP, and fetal heart NHH19W) were mixed, and ss circles
 were made in vitro. Following BAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of 1 M.A.G.E. clones 260232-265223,
 440488-445479, and 484488-489479."
 120 a 152 c 191 g 123 t 1 others
 ORIGIN
 query Match 30.7%; Score 529.2; DB 9; Length 587;
 Best Local Similarity 98.0%; Pred. No. 7.3e 72;
 Matches 578; Conservative 0; Mismatches 8; Indels 4; Gaps 4;
 1113 cagaccgaagatgaggaacattaggtgaggtgtttagctgttagcccttagctc 1172
 |||||||
 537 cagaccgaagatgaggaacattaggtgaggtgtttagctgttagcccttagctc 530
 |||||||
 1173 cctcatcctgaggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1232
 |||||||
 529 cctcatcctgaggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 470
 |||||||
 1233 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1292
 |||||||
 439 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 410
 |||||||
 1233 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1352
 |||||||
 439 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 451
 |||||||
 1333 ccaagatctcctctgaggaatctctctggtggaattctgaggtgaggtgaggtc 1412
 |||||||
 430 ccaagatctcctctgaggaatctctctggtggaattctgaggtgaggtgaggtc 291
 |||||||
 1413 tctcatcctgaggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1472
 |||||||
 230 tctcatcctgaggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 231
 |||||||
 1473 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1532

|||||
 1b 260 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 171
 |||||||
 1533 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1592
 |||||||
 170 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 111
 |||||||
 1553 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1651
 |||||||
 110 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 51
 |||||||
 1552 cactgctgctctgacaaatgaagatgattgctgctgtaaaaaa 1701
 |||||||
 50 cactgctgctctgacaaatgaagatgattgctgctgtaaaaaa 1
 |||||||
 RESULT 11
 BE439634/c 518 bp mRNA linear EST 40-MAR-2001
 LOCUS
 BE439634
 DEFINITION
 nab66c02.x1 Soares_NSF_9w_01_PA_P_S1 Homo sapiens cDNA clone
 IMAGe:4272810 3', mRNA sequence.
 ACCESSION
 BE439634
 VERSION
 BE439634.1 GI:11452072
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 518)
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps@nci.nih.gov
 This clone is available royalty free through LINC; contact the
 IMAGe Consortium (info@image.llnl.gov) for further information.
 Seq primer: 400P from G1beo
 High quality sequence stop: 454.
 FEATURES
 SOURCE
 1..518
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="Soares_NSF_9w_01_PA_P_S1"
 /issue_type="Pooled human melanocyte, fetal heart, and
 /lab_host="PH10B"
 /note="Organ: pooled; Vector: pTZ19-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following BAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares_NSHSP pool 1:
 309384-310919, 329308-329805 Soares_NSHSP pool 1:
 145032-147345, 177220-181831, 148672-149255, 15002-
 150407, 151176-152327 Soares_NSHSP pool 1:
 752489-756584, 772194-774497 Soares_NSHSP pool 1:
 304776-306811, 320136-322823, 326480-326663 Soares_NSHSP
 pool 1: 723720-726407, 736080-740999 Subtraction by hybrid
 Soares and M. Patricia Bonaldo."
 108 a 149 c 170 g 99 t 1 others
 ORIGIN
 query Match 29.9%; Score 509.8; DB 10; Length 518;
 Best Local Similarity 99.4%; Pred. No. 7.1e-69;
 Matches 511; Conservative 0; Mismatches 3; Indels 6; Gaps 0;
 1182 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 1441
 |||||||
 518 aagagaagatctggtctgacactcctctaggtcttgaacctcctcctgattgcctctg 459

• • •

[illegible]

RESDIT	6	
AA568961		
LOCUS	366 bp	linear
DEFINITION	nm2h08.s1 N1_CCAP_1P2 Homo sapiens cDNA clone IMAGE:161463	EST:09-SEP-1997
ACCESSION	AA568961	
VERSION	AA568961	
KEYWORDS	similar to contains Alu repetitive element; mRNA sequence.	
SOURCE	AA568961 1	gi:2442015
ORGANISM	HST	
	Human	
	Homo sapiens	

REFERENCE AUTHORS TITLE	JOURNAL	COMMENT
(Chapters 1 to 366) NCI-60AP data: <i>www.ncbi.nlm.nih.gov/nci60ap</i> National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
Unpublished (1997)		
Contact: Robert Strausberg, Ph.D.		

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA library preparation: David B. Kitzman, Ph.D.
cDNA library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCAP clone distribution information can be found through The L.M.A.G.E. Consortium (<http://lmi.atl.bio.tlul.dow/4btp/lmapc/lmapc.html>)
Insert length: 571 Std Error: 0.00
Seq primer: -40m13 fwd: FT from Amersham
High quality sequence stop: 554.

SOURCE	
1.	466
Accession#	"Homo sapiens"
Cell_line	"A-498"
Image	"IMAGE1061463"
Library	"NCI-CAT_112"
Tissue_type	"Liposarcoma"
Add_host	"DH10B"
Note	"vector: pAMP10; mRNA made from liposarcoma, cDNA made by oligo-dT priming. Non-directly cloned."
Size_selected	"on agarose gel, average insert size 600 bp."
Reference:	Krizman et al. (1996) Cancer Research
5615480-5183 "	
84	98
72	112

[illegible][illegible]

RESULT	7
A0752148	
LOCUS	843 bp
DEFINITION	Human brain cDNA for A0752148
ACCESSION	U00000.1
VERSION	1
KEYWORDS	Brain, cDNA
SOURCE	Human
ORGANISM	Homo sapiens

	proportion of the total population; Mammalia) Eutheria (Primates); Carnivora (Hominidae) Homo.
REFERENCE	1 (bases 1 to 843)
AUTHORS	Maitavias-G., Wallace, J.C., Smith, K., Swartzell, S., Hojman, T., Keller, A., Shaker, B., Furlong, J., Young, L., Zhou, S., Adams, M.D., and Hood, L.
TITLE	Sequence-tagged connectors: A sequence approach to mapping and assessing the human genome

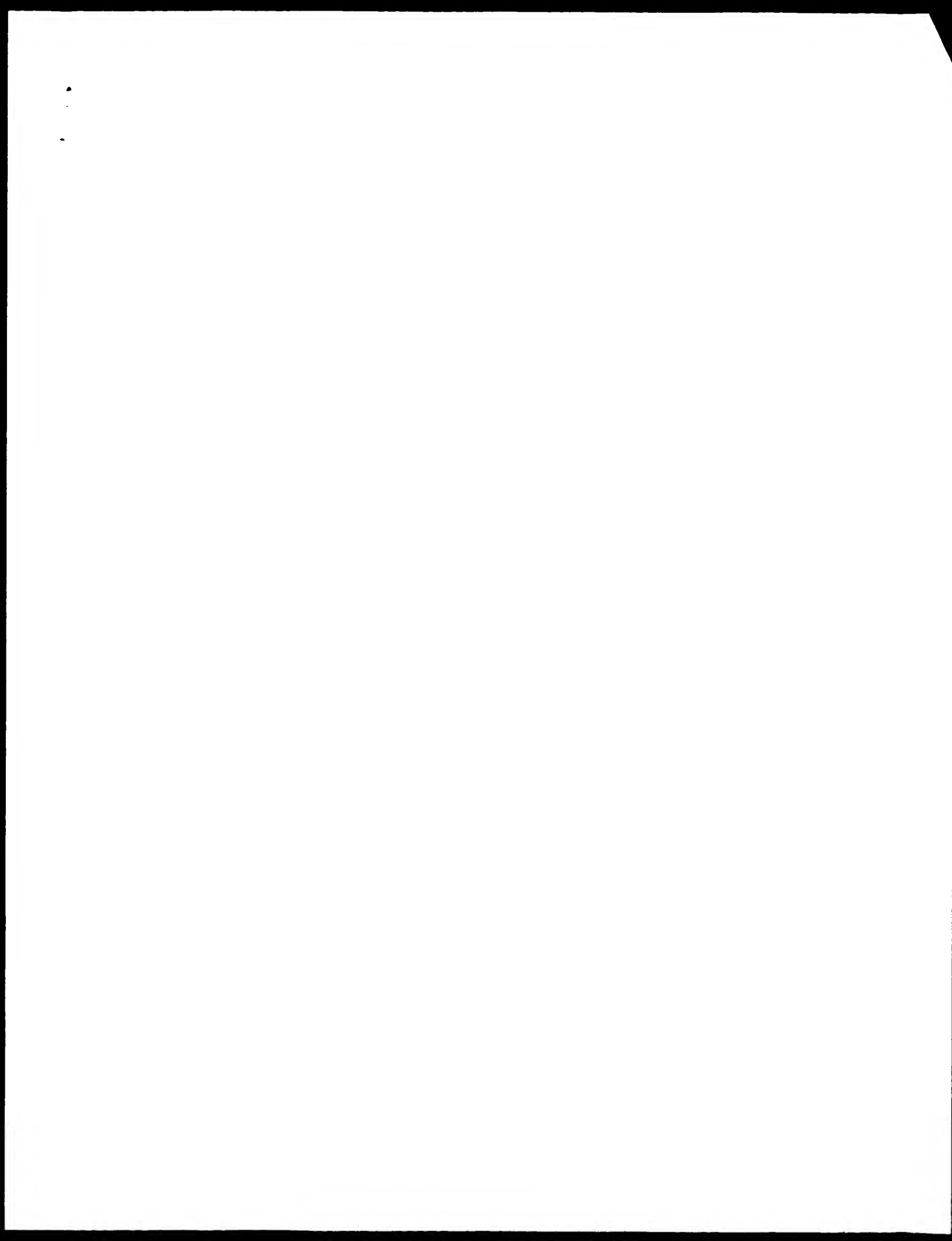
JOURNAL
OF
MEDICAL
GENETICS
COMMENT

Fluck, M., et al. (1995) 32, 334-339 (1995)
99480589

Contact: Mahabir G. Wallace MD, Board L
High Throughput Sequencing Center
University of Washington
441 Green Acme Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3618

Email: jwalace@u.washington.edu
 Clones are derived from the human *MDA* library (pgs 1-11, for MDA library availability, please contact Peter de Jong)
 (peter@dojona.med.utah.edu). Clones may be purchased from
 MDA. Accession: E141. Project: MDA. Date: 11/19/99
 or from Resear h Genetics (info@resgen.com), MDA and Web Server:
<http://www.bcr.washington.edu>
 Plate: 1147 row: H column: 4
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 843
 Location/Qualifiers
 1 843

/cloningism="vector" cat="mus"
/db_atc="taken;9605"
/clone_plate="147 Col 4 Row H"
/clone_lib="R6C1-11 Human Male, KAT library"
/sex="male"
/note="Vector: pEA2v-3'cat, 5'-cat, E10K1, E10K2, E10K3
Male biased RNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI-Methylase. Size selected DNA was cloned into the
pRCM3.6 vector at EcoRI sites"



[illegible][illegible]

[illegible][illegible]

[illegible]

01-SEP-2000; 2000US 0229343.

[illegible][illegible]

SOURCE
 ORGANISM
 Homo sapiens
 Eukaryote; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 191887)
 Birren, B., Linton, L., Nusbaum, C., and Landt, E.
 Homo sapiens chromosome 17, clone CH-2545H
 JOURNAL
 REFERENCE
 AUTHORS
 Unpublished
 2 (bases 1 to 191887)
 Birren, B., Linton, L., Nusbaum, C., Landt, E., Allen, N.,
 Anderson, S., Patton, N., Bastien, V., Beaulieu, J., Beaulieu, B.,
 Brown, A., Camarero, J., Campiano, A., Chang, J., Chazotte, B.,
 Choquet, Y., Colangelo, M., Collins, S., Collinge, A., Cook, A.,
 Cooke, P., Davidson, K., Dewar, K., Dhanraj, S., Dhanraj, S.,
 Ferris, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
 Ginde, S., Gird, S., Goyette, M., Graham, L., Grandpierre, N.,
 Hago, B., Heald, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kawai, A., Karatas, A., Kelly, C., Laporte, K.,
 Lamas, R., Landt, E., Lebeck, J., Levine, R., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McKen, P., McKernan, K., McPherson, R., McPherson, T.,
 Meneses, L., Milnova, T., Mienda, V., Murphy, T., Naylor, J., Nguyen, C.,
 Nord, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, T., Peterson, K., Phukhant, P., Pierre, N., Pollard, V.,
 Raymond, G., Rella, F., Ribick, M., Riley, R., Rise, C., Rong, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schaefer, S., Schpack, R.,
 Seaman, S., Sever, P., Spencer, H., Stagg, P., Stagg, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
 Topfman, K., Travers, M., Travis, N., Trifoglio, J., Vassiliev, R.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, C.,
 Zaidman, J., Zambek, L., Zimmer, A., and Zody, M.
 Direct Submission
 Submitted (Jan-Mar-2001) to the GenBank database for the GenBank
 Research, 420 Charles Street, Cambridge, MA 02141, USA
 All reports were identified using the keyword: "Homo sapiens"
 Smith, A.F.A. & Green, P. (1996-1997)
 http://trf.biology.oxford.ac.uk/bp/bp.html
 TITLE
 JOURNAL
 COMMENT
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WILBR
 Web site: <http://www.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project information
 Center project name: 110708
 Center clone name: 2545.H.1
 Summary Statistics
 Sequencing vector: pUC19, n/7, 100% of reads
 Chemistry: dye-terminator big dye, 100% of reads
 Assembly program: Phrap, version 0.960731
 Consensus quality: 182941 bases at least 400
 Consensus quality: 187111 bases at least 400
 Consensus quality: 188492 bases at least 400
 Insert size: 195000; average=19
 Insert size: 189187; sum-of-counts
 Quality coverage: 11.7 in 920 bases; sum of counts
 Quality coverage: 12.0 in 920 bases; sum of counts
 NOTE: This is a working draft sequence. It currently
 consists of 96 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1
 643 742: gap of 100 bp
 743 1333: contig of 591 bp in length
 1334 1443: gap of 100 bp
 1444 2386: contig of 903 bp in length
 2387 2437: gap of 100 bp
 2438 3175: contig of 737 bp in length
 3176 4275: gap of 100 bp

FEATURES
 SOURCE
 misc_feature
 3276 4422: contig of 1147 bp in length
 4423 4522: gap of 100 bp
 4523 5142: contig of 620 bp in length
 5143 5242: gap of 100 bp
 5243 5908: contig of 666 bp in length
 5909 6008: gap of 100 bp
 6009 6926: contig of 918 bp in length
 6927 7026: gap of 100 bp
 7027 7674: contig of 648 bp in length
 7675 7774: gap of 100 bp
 7775 8461: contig of 587 bp in length
 8462 9486: gap of 100 bp
 9487 9186: gap of 100 bp
 9187 10016: contig of 830 bp in length
 10017 10116: gap of 100 bp
 10117 10729: contig of 613 bp in length
 10730 10929: gap of 100 bp
 10930 11973: contig of 1144 bp in length
 11974 12073: gap of 100 bp
 12074 14092: contig of 1019 bp in length
 14093 13192: gap of 100 bp
 13193 14041: contig of 839 bp in length
 14042 14131: gap of 100 bp
 14132 14841: contig of 700 bp in length
 14832 14931: gap of 100 bp
 14932 16704: gap of 1672 bp in length
 16595 16694: gap of 100 bp
 16695 17768: contig of 1074 bp in length
 17769 17868: gap of 100 bp
 17869 18647: contig of 769 bp in length
 18638 18737: gap of 100 bp
 18738 20094: contig of 1259 bp in length
 20037 20136: gap of 100 bp
 20137 22573: contig of 2136 bp in length
 22574 22723: gap of 100 bp
 22724 22773: gap of 100 bp
 22774 23931: contig of 1158 bp in length
 23932 23931: gap of 100 bp
 23932 25353: contig of 1423 bp in length
 25354 25453: gap of 100 bp
 25454 34669: contig of 9216 bp in length
 34670 34769: gap of 100 bp
 34770 79954: contig of 45185 bp in length
 79955 80054: gap of 100 bp
 80055 139801: contig of 59747 bp in length
 139802 139901: gap of 100 bp
 139902 191887: contig of 51886 bp in length.
 local/contig/Qualifiers
 1. 191887
 /organism="Homo sapiens"
 /full_name="Homo sapiens"
 /chromosome="17"
 /map="17"
 /clone="CH-2545H"
 /clone_11b="CH-2545H Human BAC"
 1. 642
 /note="assembly-fragment"
 743. 1333
 /note="assembly-fragment"
 1334. 2386
 /note="assembly-fragment"
 2387. 3175
 /note="assembly-fragment"
 3176. 4275
 /note="assembly-fragment"
 4276. 5142
 /note="assembly-fragment"
 5143. 5908
 /note="assembly-fragment"
 5909. 6926
 /note="assembly-fragment"
 6927. 7674
 /note="assembly-fragment"

putting sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/Sceleropterus/wormpep>. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/chr6/>.

Rp11-54923 is from the library RP11-11.2 constructed at the Roswell Park Cancer Institute by the group of Pierre de Jong. For further details see <http://www.rpi.ccr.tycho.edu/>.

IMPORTANT: This sequence is not the entire insert of clone Rp11-54923. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone Rp11-54923 is at 18547 in this sequence. The true left end of clone Rp11-44805 is at 86817 in this sequence. The true right end of clone Rp3-32572 is at 100 in this sequence.

FEATURES

```

SOURCE
    1..185467
    /formation: "Homo sapiens"
    /db_xref: "taxon 9606"
    /chromosome: "6"
    /clone: "Rp11-54923"
    /clone: "11b-Rp11-11.2"
    1..204
    repeat_region
        /note: "Alu repeat: matches 97..204 of consensus"
        /note: "match: SSS: Em:AG267967"
        376..852
        misc_feature
            /note: "match: SSS: Em:AG267967"
            989..1276
        repeat_region
            /note: "Alu repeat: matches 1..287 of consensus"
            2200..4475
        repeat_region
            /note: "Alu repeat: matches 347..411 of consensus"
            4973..5266
        repeat_region
            /note: "Alu repeat: matches 1..294 of consensus"
            5431..5807
        repeat_region
            /note: "MIR repeat: matches 1..390 of consensus"
            5907..6258
        misc_feature
            /note: "match: SSS: Em:AG788293"
            6548..6686
        repeat_region
            /note: "LIM4 repeat: matches 5236..5794 of consensus"
            6735..7223
        repeat_region
            /note: "LIM4 repeat: matches 1472..2061 of consensus"
            7284..7580
        repeat_region
            /note: "Alu repeat: matches 1..298 of consensus"
            7581..7986
        repeat_region
            /note: "LIM4 repeat: matches 1866..1462 of consensus"
            8018..8159
        repeat_region
            /note: "LIM4 repeat: matches 2..141 of consensus"
            8303..8588
        repeat_region
            /note: "LIM4 repeat: matches 449..759 of consensus"
            8712..8928
        repeat_region
            /note: "Alu repeat: matches 85..300 of consensus"
            complement(9457..9648)
        misc_feature
            /note: "match: SSS: Em:AG353367"
            9668..9967
        repeat_region
            /note: "Alu repeat: matches 6..307 of consensus"
            10859..11098
        repeat_region
            /note: "LIM4 repeat: matches 5855..6680 of consensus"

```

```

repeat_region
    11099..11394
    /note: "Alu repeat: matches 1..294 of consensus"
    13689..13998
    repeat_region
        /note: "Alu repeat: matches 1..411 of consensus"
        15065..15567
    repeat_region
        /note: "L2 repeat: matches 1810..2455 of consensus"
        15767..16026
    repeat_region
        /note: "Alu repeat: matches 1..289 of consensus"
        complement(15863..16438)
    misc_feature
        /note: "match: SSS: Em:AG577248"
        16029..16086
    repeat_region
        /note: "LIM4 repeat: matches 5797..5942 of consensus"
        17041..17044
    misc_feature
        /note: "match: SSS: Em:AG072859"
        17640..17991
    repeat_region
        /note: "match: SSS: Em:AG3627"
        18647..18949
    repeat_region
        /note: "Alu repeat: matches 1..404 of consensus"
        complement(19574..20078)
    repeat_region
        /note: "match: SSS: Em:AG777121"
        20434..20800
    repeat_region
        /note: "Alu repeat: matches 1..497 of consensus"
        21664..21560
    repeat_region
        /note: "Alu repeat: matches 1..287 of consensus"
        21931..22230
    repeat_region
        /note: "Alu repeat: matches 1..295 of consensus"
        22577..22356
    repeat_region
        /note: "MIR repeat: matches 155..260 of consensus"
        22821..23119
    repeat_region
        /note: "Alu repeat: matches 2..292 of consensus"
        23780..23903
    repeat_region
        /note: "LIM4 repeat: matches 5712..5886 of consensus"
        24212..24510
    repeat_region
        /note: "Alu repeat: matches 1..297 of consensus"
        25084..25216
    repeat_region
        /note: "L2 repeat: matches 3779..2710 of consensus"
        25749..25903
    repeat_region
        /note: "PRAM repeat: matches 9..163 of consensus"
        complement(26457..27119)
    misc_feature
        /note: "match: SSS: Em:AF191939"
        27133..27822
    misc_feature
        /note: "match: SSS: Em:AF191939"
        27319..27648
    repeat_region
        /note: "Alu repeat: matches 1..312 of consensus"
        28191..28418
    repeat_region
        /note: "Alu repeat: matches 1..298 of consensus"
        28719..28978
    repeat_region
        /note: "LIM4 repeat: matches 1..464 of consensus"
        30450..30741
    repeat_region
        /note: "LIM4 repeat: matches 5574..5865 of consensus"
        31560..31703
    repeat_region
        /note: "LIM4 repeat: matches 5797..5942 of consensus"
        31704..32012
    repeat_region
        /note: "Alu repeat: matches 1..412 of consensus"
        32033..32346
    repeat_region
        /note: "LIM4 repeat: matches 7428..5797 of consensus"
        32831..33017
    repeat_region
        /note: "Alu repeat: matches 1..250 of consensus"
        33018..33048
    repeat_region
        /note: "Alu repeat: matches 218..248 of consensus"
        34161..34432
    repeat_region
        /note: "Alu repeat: matches 25..304 of consensus"
        34468..34514
    repeat_region
        /note: "MIR repeat: matches 196..247 of consensus"
        34854..35144
    repeat_region
        /note: "Alu repeat: matches 6..276 of consensus"
        35394..35445
    repeat_region
        /note: "MIR repeat: matches 161..250 of consensus"
        35446..35489
    repeat_region
        /note: "L2 repeat: matches 2653..2705 of consensus"
        36738..37043
    repeat_region
        /note: "Alu repeat: matches 1..295 of consensus"
        37167..37248

```



```

COMMENT      on Aug 20, 2001 this sequence version replaced 4114572572.
FEATURES
    source
        location/qualifiers
            1..104429
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="8"
                /map="8q24"
                /clone="K1404011"
                /cell_line="FEB14 - 14"
                /clone_lib="Koto BMC library"
                239..1311
                    /rpt_family="L1ME8"
                    /evidence=not_experimental
                    1322..1503
                        /rpt_family="L1"
                        /evidence=not_experimental
                    1532..1645
                        /rpt_family="L1"
                        /evidence=not_experimental
                    1636..2236
                        /rpt_family="L1ME8"
                        /evidence=not_experimental
                    3188..3625
                        /rpt_family="MER83"
                        /evidence=not_experimental
                    3815..3846
                        /rpt_family="AL_rich"
                        /evidence=not_experimental
                    3987..3989
                        /rpt_family="L1MA7"
                        /evidence=not_experimental
                    4015..4066
                        /rpt_family="CT_rich"
                        /evidence=not_experimental
                    4254..5394
                        /rpt_family="L2"
                        /evidence=not_experimental
                    5997..6447
                        /rpt_family="M1110"
                        /evidence=not_experimental
                    6450..6751
                        /rpt_family="A1"
                        /evidence=not_experimental
                    6193..8422
                        /rpt_family="M18"
                        /evidence=not_experimental
                    8673..8695
                        /rpt_family="(1)n"
                        /evidence=not_experimental
                    9425..9477
                        /rpt_family="L2"
                        /evidence=not_experimental
                    11083..11168
                        /rpt_family="L2"
                        /evidence=not_experimental
                    12457..13157
                        /rpt_family="L1MR7"
                        /evidence=not_experimental
                    13383..13499
                        /rpt_family="M18"
                        /evidence=not_experimental
                    13641..13756
                        /rpt_family="M11"
                        /evidence=not_experimental
                    14172..14241
                        /rpt_family="M18"
                        /evidence=not_experimental
                    14572572

```

```

repeat_region complement(14242..14562)
                /rpt_family="A1"
                /evidence=not_experimental
                complement(14563..14603)
                    /rpt_family="M18"
                    /evidence=not_experimental
                    complement(15104..15281)
                        /rpt_family="M18"
                        /evidence=not_experimental
                    15282..15344
                        /rpt_family="AT_rich"
                        /evidence=not_experimental
                    16381..16505
                        /rpt_family="M18"
                        /evidence=not_experimental
                    17632..17652
                        /rpt_family="L1R16A"
                        /evidence=not_experimental
                    17802..17976
                        /rpt_family="M18"
                        /evidence=not_experimental
                    18059..18466
                        /rpt_family="M18A"
                        /evidence=not_experimental
                    18560..18753
                        /rpt_family="M11K"
                        /evidence=not_experimental
                    20182..20182
                        /rpt_family="MER2"
                        /evidence=not_experimental
                    20141
                        /rpt_family="(10)n"
                        /evidence=not_experimental
                    20167..20204
                        /rpt_family="(CA)n"
                        /evidence=not_experimental
                    20957..20996
                        /rpt_family="AT_rich"
                        /evidence=not_experimental
                    21506..21581
                        /rpt_family="MADE1"
                        /evidence=not_experimental
                    22892..22899
                        /rpt_family="L10B1"
                        /evidence=not_experimental
                    22437..22479
                        /rpt_family="(1A)n"
                        /evidence=not_experimental
                    24124..24665
                        /rpt_family="M18"
                        /evidence=not_experimental
                    25313..25389
                        /rpt_family="MER34"
                        /evidence=not_experimental
                    25390..25694
                        /rpt_family="7S1RNA"
                        /evidence=not_experimental
                    25694..25728
                        /rpt_family="(CA1A)n"
                        /evidence=not_experimental
                    25732..25967
                        /rpt_family="MER34"
                        /evidence=not_experimental
                    26119..26405
                        /rpt_family="A1"
                        /evidence=not_experimental
                    26478..26565
                        /rpt_family="MER58"
                        /evidence=not_experimental
                    26589..26709
                        /rpt_family="L2"
                        /evidence=not_experimental
                    27044..27124

```